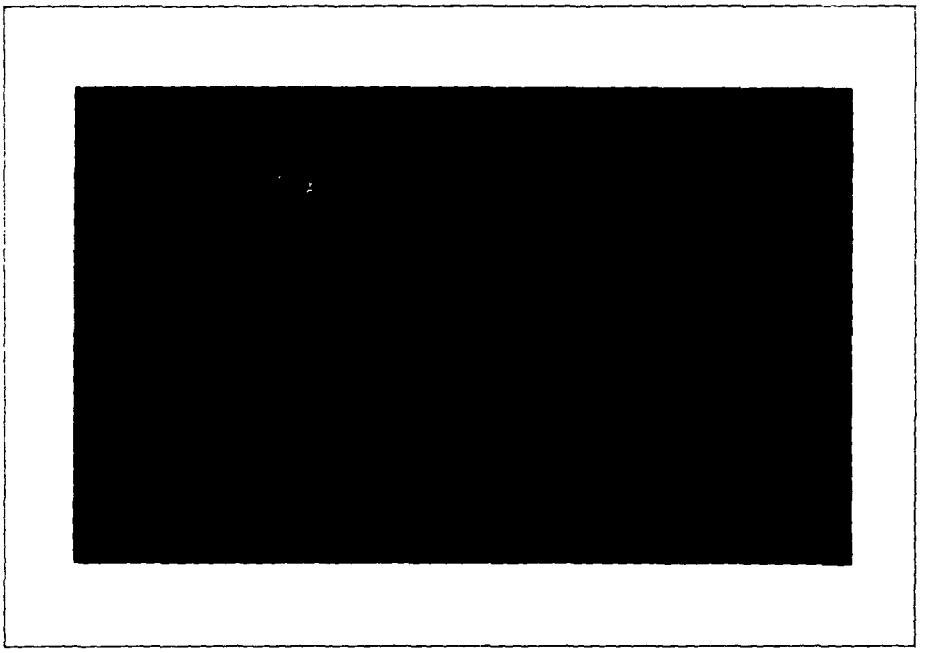

The Artificial Intelligence and Psychology Project

Departments of
Computer Science and Psychology
Carnegie Mellon University

Learning Research and Development Center
University of Pittsburgh

IMAGES OF EMPTY SPACE: EINSTEIN'S WORD PICTURES

Technical Report AIP - 90

Herbert A. Simon

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

10 August 1988

This research was supported by the Computer Sciences Division, Office of Naval Research, under contract number N00014-86-K-0678. Reproduction in whole or part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Field	Entry
1a. REPORT SECURITY CLASSIFICATION	Unclassified
1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
3. DISTRIBUTION / AVAILABILITY OF REPORT	Approved for public release; Distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	AIP - 90
5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION	Carnegie-Mellon University
6b. OFFICE SYMBOL (If applicable)	
6c. ADDRESS (City, State, and ZIP Code)	Department of Psychology, Pittsburgh, Pennsylvania 15213
7a. NAME OF MONITORING ORGANIZATION	Computer Sciences Division, Office of Naval Research
7b. ADDRESS (City, State, and ZIP Code)	800 N. Quincy Street, Arlington, Virginia 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION	Same as Monitoring Organization
8b. OFFICE SYMBOL (If applicable)	
8c. ADDRESS (City, State, and ZIP Code)	
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	N00014-86-K-0678
10. SOURCE OF FUNDING NUMBERS	p4000ub201/7-4-86

PROGRAM ELEMENT NO	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
N/A	N/A	N/A	N/A

11. TITLE (Include Security Classification)

Images of Empty Space: Einstein's Word Pictures

12. PERSONAL AUTHOR(S) Herbert A. Simon

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	1988 August 10	18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

problem solving, problem representation, diagrams, visual imagery

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

SEE REVERSE SIDE

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT

☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. Alan L. Meyrowitz	(202) 696-4302	N00014

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

ABSTRACT

Einstein can hardly be called prolix in his initial presentation of the Theory of Relativity. Nor can we suppose that he was writing down to his readers. Hence the relative simplicity of the images he asks his readers to construct is striking.

The images of the paper on relativity can perhaps provide us with at least some estimates of the upper bounds of human imaging capability -- of the extent to which problems have to be factored into their component parts before the human mind can encompass them.

Cont on Page 3
1473

Images of Empty Space: Einstein's Word Pictures

Herbert A. Simon
Carnegie-Mellon University

The claim is often repeated in the popular press that only twelve people understand the theory of relativity. The number did not seem to change at the death of Einstein, though one would have thought that it would have decreased to eleven. No matter: the claim is patently false. The theory of special relativity is now commonly taught as part of first-year college physics courses, with the mathematical formulas included; and more or less bowdlerized versions of the general theory, with no attempt at the mathematics, are also often to be found in the textbooks for such courses. There are in the population some tens of thousands of people who have taken and passed these courses -- hence who can be presumed to understand at least special relativity.

But do I now not err on the other side? What can we infer from a student's passing a physics course, or even passing an examination on special relativity theory, about his or her understanding of the subject? And what would we accept as evidence for such understanding? It is the purpose of this paper to explore the meaning of the term "understanding," using special relativity theory as a vehicle for the exploration.

Einstein's 1905 Relativity Paper

But even this is too large a topic, and it must be narrowed further. This essay will be limited to examining what it might mean for a reader to understand the 1905 paper, *On the Electrodynamics of Moving Bodies*, in which Einstein first set forth the theory of special relativity.[1] And we will consider only the first, kinematic, portion of the paper (in fact, only the first three sections, totaling 11 pages), leaving aside the applications and the discussion of electrodynamics.

Einstein's exposition has an air of simplicity, that I think most readers ultimately find deceptive. The natural language of the text is lucid and direct, full of examples that are quite

[1] I will use the well-known English translation of the paper that is to be found in *The Principle of Relativity*, Dover Publications, 1923.

concrete. The mathematics is elementary: it involves only algebra, except for one step that requires some simple calculus (and that could have been carried out algebraically). There are no figures in the entire paper.

Outline of Einstein's Paper

The argument proceeds as follows: In an introduction of a little less than two pages, Einstein mentions some difficulties, theoretical and experimental, in current physical theories relating to the concepts of absolute rest and of the luminiferous ether. He then briefly states a "principle of relativity," which, he says, will remove the difficulties he has pointed to in Maxwell's theory of electrodynamics. He claims, further, that these difficulties derive from inadequate attention to the underlying kinematics of rigid bodies.

Einstein is now ready for the kinematical analysis. In Section 1, a little more than two pages, he proposes a careful, operational definition of the simultaneity of two clocks that are both at rest in the same reference frame. In Section 2, also two pages, he applies the definition to show that two clocks that are synchronized in one reference frame would not appear synchronous to observers who are in motion relative to that frame. Up to this point, only four rather trivial algebraic equations have been displayed, each carefully motivated by the natural language text.

This demonstration sets the stage for Section 3 (5 pages), where the equations of special relativity for the transformation of coordinates and times from one reference frame to another are derived in full. The first page describes the physical situation to which the transformation is to be applied: two systems of coordinates, one regarded as stationary, the other moving, relative to the first, with constant velocity along the X axis, and each supplied with clocks and rigid measuring rods that move with it.

The main part of the derivation occupies the next two pages. Another page is devoted to showing that the result is compatible with the assumption of the constancy of the speed of light. The final page is given over to showing that an unknown function, appearing in the transformations as initially derived, is equal to unity.

Nature of the Explanation

The explanation of the theory of relativity that Einstein provides in these pages involves, as do all logical arguments, some assumptions and some conclusions derived from them. The derivations are mathematical, employing relatively simple algebraic manipulations (although some intermediate steps are omitted). Anyone having even rather modest mathematical competence can verify that these steps of derivation are valid -- that each follows from its predecessors. To the extent that "understanding" means being able to follow a proof and see its correctness, understanding relativity does not appear to be too difficult a task. Clearly, other things are involved.

Every mathematical derivation must begin with premises: equations that are not themselves derived. The principal premises for Einstein's algebraic argument appear on the second page of Section 3 (page 44 of the Dover edition). They are three in number: a definition, introducing a symbol, x', which simplifies the algebra; an equation relating the times and positions of clocks that are synchronized in reference frames that are in motion relative to each other; and an assertion that a certain velocity has a particular value. On the next page, equations are introduced to express the fact that the coordinates of a ray of light, measured in relation to a particular reference frame and passing through the origin at time zero, would be proportional to the product of the speed of light and the time (measured in that reference frame). All other equations on these two pages are derived from these.

But why these premises instead of others? Most of the prose is devoted to motivating them. Hence, one may assume that part of understanding Einstein's argument consists in understanding these premises and finding convincing reasons in the natural language text for accepting them. But this reasoning is not formalized. We must ask how a reader might check it.

Another part of understanding the argument must involve ascribing a physical "meaning" to the equations that are derived (perhaps not only the final equations but also some of those that appear in the course of the derivation). Section 4 of Einstein's paper, also two pages long, is devoted to interpretation of the physical implications of the

transformations, but there is relatively little such interpretation in the pages on which we are focusing. However, the reader might be expected himself or herself to make such interpretations while trying to understand the argument and its result.

But our first task is to understand the premises of Einstein's algebraic derivation. To do that, we must turn to the natural language prose that introduces and motivates his initial equations.

The Basic Assumptions

Einstein relies on two explicit assumptions -- the Principle of Relativity and the constancy of the speed of light. He introduces these two assumptions, informally, in his introduction (top of page 38), but defines them more carefully at the beginning of Section 2 (page 41).

> 1. [Principle of Relativity] The laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to one or the other of two systems of coordinates in uniform translatory motion.
> 2. [Speed of light] Any ray of light moves in [a] "stationary" system of coordinates with the determined velocity c, whether the ray be emitted by a stationary or by a moving body.

These postulates would probably evoke, in a physicist of 1905, some of the following thoughts: the principle of relativity was already a part of Newtonian mechanics, whose laws (as Newton observed) were invariant under Galilean (uniform, unaccelerated) transformations, but not under acceleration or rotation of the reference frame. However, as Einstein observed (and as was well known to physicists), Maxwell's Laws, which governed electromagnetic phenomena were *not* invariant under Galilean transformations.

What Einstein promises, then, is a reformulation of the laws of physics that will make both the laws of mechanics and the laws of electrodynamics invariant under uniform translations. (The physicist would also be aware of other anomalies, all relating to the speed of light, that made difficulties for the principle of relativity in classical physics, and Einstein refers to these elliptically in his introduction.)

In his introduction, Einstein states his postulate of the constancy of the speed of light a little differently from the way it is stated above. In the introduction he refers (page 38) to:

> another postulate, which is only apparently irreconcilable with the [Principle of Relativity], namely, that light is always propagated in empty space with a definite

velocity *c* which is independent of the state of motion of the emitting body.

The reference to "apparent irreconcilability," not repeated on page 41, is interesting. Interpreting it can serve as our first exercise in explicating what it means to "understand" Einstein's paper. How would the reader have to understand the emission of light in order to find the constancy of the speed of light irreconcilable with the principle of relativity? A physicist committed to the belief that light is transmitted through a stationary ether (as many, or perhaps most, physicists of 1905 were) would also believe that, relative to a given reference frame, a ray of light would have the same velocity whether emitted from a stationary or a moving body, because the light is moving in the ether.

But that same physicist would observe that if he measured the speed of light in a different reference frame, one that moved relative to the ether, the measured speed would be different. Thus the principle of relativity would be violated for Galilean transformations of the reference frame. It was this consideration that motivated the famous Michelson-Morley experiments.

An alternative version of the ether theory, which was also sometimes entertained, would have each massive body carry along with it the surrounding ether. Then, from an external point of observation, the velocity of light emitted along the axis of motion of a body moving relative to that external reference frame would, for the observer, be different from the velocity of light emitted by a body that was stationary in that external frame. In this case, the Principle of Relativity holds, but not the constancy of the speed of light.

The three preceding paragraphs are, as we have said, an exercise in understanding. We have asked the reader to understand *why* Einstein thought two principles were "apparently irreconcilable," and we have proposed that the reader achieve understanding by performing some thought experiments.

The first thought experiment requires visualizing a space occupied by a stationary substance (the ether). Light is then transmitted through this substance, always with the same velocity. Now it can be "seen" that if we visualize ourselves as moving relative to that substance, and measure the speed of light in relation to our own co-ordinates, the speed will

no longer be uniform, but will depend on our motion relative to the ether. The second thought experiment, with an ether that moves locally with each body, again allows us to "see" that the two principles contradict each other.

Understanding by Imaging

Understanding, it would seem (at least in situations like these), has something to do with "seeing" in the mind's eye. It is well known that most, perhaps all, people have capabilities for constructing mental representations of spatial scenes, and -- most important -- have capabilities for drawing inferences from those visualized scenes. As a simpler example than the one before us, we may ask a subject to visualize a rectangle, and then to add to the image its two diagonals. Now the subject, asked whether the diagonals have a point of intersection, will reply that they do. The point of intersection was not explicitly given; it was inferred by the act of visualization.

There is nothing magical or mysterious about such inferencing. Even if we do not know specifically how it is done in the brain, we do know how it is done by much simpler systems. For if we draw the rectangle on a piece of paper, together with its two diagonals, we also record the point of intersection "for free." The act of drawing on a piece of paper is evidently also an inferencing process, and a very powerful one.

For our purposes here, we will simply assume that visualizing a situation in the mind's eye causes inferences to be made that are quite analogous to those that are made by drawing on paper. When we "see" the mental picture of the rectangle with its diagonals and intersection, we acknowledge that we understand why the intersection is present. We simply cannot image the diagonals without imaging the intersections.

What the mind's eye sees need not be simply a static picture. At least to some extent, motion can also be visualized. What can this mean? Let us again invoke the aid of a drawing on paper, the paper now representing a fixed reference frame. Place a dot on the paper to designate a fixed point, and label it A. Place a second dot to the right of the first, and label it B. Now, consider a ray of light that starts from A at time t_A, moving toward B. At time t_B it is reflected from B and moves back to A, arriving at a time we will call t'_A. The movement can be

visualized by considering a point that begins at A, and at each successive moment has a new position further to the right until it reaches B. Then it takes on successive positions, each further to the left, until it returns to A.

There are a number of ways in which such a succession of images could be realized in a "mind's eye." The moving point suggested above corresponds to the simplest way we could represent the movement on a computer screen. In a textbook figure, it might be represented, less adequately, by an arrow. Psychological exeriments on apparent motion suggest that the human brain has other, specialized, mechanisms for perceiving stimuli as in motion. Just how the mind does it is not important to us. What are important is the fact that it can be done, and the fact that, if we do not know exactly how it is done by this particular mechanism, we do know how it could be done by other mechanisms.[2]

Nor, for our purposes, is it important whether a mental image can provide information about continuous motion. In Einstein's relativity paper, the only motions considered are movements of a point from one position to another and movements of a light ray from one point to another. Visualizing these motions in order to understand the equations does not require a moving picture of the events, but only static before-and-after snapshots. This is true for our simple example, above, of a light ray that follows a path from one point to another and back, and it is equally true for all of the other examples we shall have to consider.

On the basis of these considerations, we can now state our general thesis and proceed to test it. *The general thesis is that Einstein motivates each of his initial equations by inviting the reader to form a mental picture that exemplifies it. The equations can, so to speak, be read directly from these mental pictures, so that when the reader has accomplished this, he or she understands the corresponding equation.*

We shall examine this hypothesis by applying it to the first three sections of Einstein's paper.

---

[2]The moving target indicator used in radar to screen out the background images of stationary objects provides another suggestion of how the eye, by substraction of successive signals, can discriminate moving stimuli from static ones.

Definition of Simultaneity

We have already introduced the picture that Einstein uses to define simultaneity. It can be found in his article beginning at the last paragraph of page 39. We have a single space, in which two points are singled out. We are asked to imagine a ray of light going from the first point to the second, and reflected back to the first. We assume that a clock is placed at each point, and ask how we would test whether the clocks are synchronized.

As before, let t_A, t_B, and t'_A be the starting time, the time at reflection, and the return time of the light ray, measured in each case on the clock local to the event. The assumption of the constancy of the speed of light, and the fact that the same distance is traversed in each case, allows us to conclude that the time from emission to reflection must equal the time from reflection to return to the starting point. Therefore, if the clocks are synchronous, we must have:

$$t_B - t_A = t'_A - t_B$$

"Thus," says Einstein (page 40), "with the help of certain imaginary physical experiments we have settled what is to be understood by synchronous stationary clocks located at different places. . . The 'time' of an event is that which is given simultaneously with the event by a stationary clock located at the place of the event, this clock being synchronous, and indeed synchronous for all time determinations, with a specified stationary clock."

Relativity of Time to Reference Frame

Showing that the synchronization of clocks is dependent on the movement of the reference frame from which they are observed is the main goal of Section 2 of Einstein's paper. Again, the product of the analysis consists of two simple equations that are motivated by a mental picture. The picture, however, is somewhat more complex than the one used in the previous section, and Einstein's description of it is not as perspicuous. I shall describe it, therefore, a little more explicitly than he did.

This time we are invited to imagine two reference frames, or systems of co-ordinates, whose X-axes coincide, and whose Y-axes and Z-axes are parallel, respectively. The first

reference frame is designated as "stationary" -- presumably relative to the vantage point from which we are viewing it. The second moves along the X-axis, relative to the first, at a uniform velocity, *v*.

We again select two points, A and B, which move with the second (moving) reference frame, and associate with each a clock that is synchronized at all times with the clock at the corresponding location, at that time, in the stationary reference frame. As before, a light ray traverses a path from A to B and returns to A. We observe the times (measured on the specially synchronized clocks at A and B) at which the ray leaves A, reaches B, and returns to A. This is equivalent to fixing the times of these three events in the moving system from the vantage point of the stationary system.

Since (still viewing the situation from the stationary system) the ray is moving towards B with a velocity of *c* while B is moving away from the original position of A (the position when the ray was emitted) with a velocity of *v*, the net velocity of the ray toward the moving B is $(c - v)$. Applying the corresponding argument to the return trip, the velocity of approach on this leg of the journey is $(c + v)$. Take the times of the three events as defined above, and call the length of the rod as observed in the stationary system *d*. Then the time it takes the ray to go from A to B will be $d/(c - v)$, while the time it takes the ray to return from B to A will be only $d/(c + v)$. This is consistent with the fact that the first path (because of the movements of A and B) is longer than the second.

Now we imagine that the same three events are viewed by observers who move with the moving reference frame, but who measure time by the clocks at A and B that are continually synchronized with clocks in the stationary frame. The rod, of length *d'* (which may or may not be equal to *d*) does not move relative to the moving frame, hence the path of the ray from A to B is the same length as the path from B back to A. Since the velocity of light is assumed to be *c* for observers in any reference frame, then if the time were measured by clocks synchronized in the moving reference frame, we should have $t_A - t_B = t'_A - t_B$. Since this relation does not hold for the actual clocks the moving observers have with them, which were synchronized relative to the stationary frame, they decide that these clocks are not

synchronous.

Einstein concludes:

> So we see that we cannot attach any *absolute* significance to the concept of simultaneity, but that two events which viewed from a system of co-ordinates, are simultaneous, can no longer be looked upon as simultaneous events when envisaged from a system which is in motion relatively to that system.

Analogous Problems in Elementary Algebra

Before we continue with Einstein's derivation of the Lorentz transformations, we wish to draw attention to a striking and exact analogy between the thought experiment we have just performed and one that we were confronted with in our high school days. In trying to form the mental pictures discussed in the last section, we may well have been reminded of the problems, beloved of textbooks in elementary algebra, about boats that travel up and down rivers, and airplanes that fly with and against the wind. A typical problem of this kind reads:

> A boat travels upstream on a river from A to B, then returns to A. The speed of the boat in still water is 12 miles per hour, and the river has a current of 4 miles an hour. The round trip takes 6 hours. What is the distance from A to B?

If we let d be the distance between A and B, v the speed of the boat, and c the speed of the current, then the upstream time is $d/(v-c)$ while the downstream time is $d/(v+c)$. The total time is the sum of these two quantities, and since this is known, as well as v and c, the equation can be solved for d.

But these times -- with a change in sign -- are precisely the times that we obtain in Einstein's second thought experiment, where the light source is moving relative to the observer's reference frame.

What kind of a thought experiment are we asking the algebra student to perform to see the correctness of the river boat equations? We ask the student to stand on the river bank, and observe that the speed of the current is always c relative to his or her reference frame (hence also relative to A and B). Now we point out that the speed of the boat *relative* to the river current is v, no matter which direction it is going. Therefore, we can see that, relative to the bank, it is $v+c$ when the boat is going downstream and $v-c$ when the boat is going upstream.

In the river problem, c and v have interchanged their roles, as compared with the roles

of these same symbols in the light-beam problem. That accounts for the change in sign. In the river problem, we have asked the student to take a reference frame that is fixed with respect to A and B, while in the relativity problem, the reader is asked to view the situation from a frame that is moving relative to A and B. In all other respects, the problems are the same. It appears that understanding river problems requires a thought experiment just as difficult as the one needed to understand special relativity.

What happens if we alter the mental picture for the river problem to match exactly Einstein's second thought experiment? Now, the observer should be on a raft, riding with the river current, so that the river is stationary in the observer's frame of reference. The points A and B have a velocity, $-c$ relative to the observer on the raft, and the boat, with a velocity, v in that same reference frame, takes the role of the light ray, "emitted" at point A, and "reflected" at point B.

Although the equations we can now write down are the same as before, intuitively, this picture seems more difficult to form than the usual picture of the river problem. But perhaps this is because we normally think of solid land, rather than a floating raft, as the appropriate basis for defining a fixed reference frame.

What happens if we try to formulate Einstein's thought experiment in the usual form of the river problem -- that is with the river bank and the points A and B as the fixed reference frame instead of the raft on the flowing water? Now the boat can no longer serve as the analogue of the light wave, because its velocity is not constant relative to the reference frame, but depends on whether it is moving upstream or down. Hence, although, as we have seen we can compute correctly the upstream and downstream speeds of the boat, there is nothing in this mapping that corresponds to the relativistic assumption that the speed of light is independent of the reference frame from which the light is observed.

Whether our intuitive feeling is valid -- that the thought experiment for special relativity is more difficult than the one for the river problem -- is an empirical question to be decided by experiment. Until the experiment is performed, we must keep our minds open. It would be remarkable indeed if the critical thought experiment that is needed to understand special

relativity turned out to be no harder than an analogous experiment that every high school student is supposed to be able to perform.

The Lorentz Transformations

We return now to the topic of Einstein's derivation of the Lorentz transformations. Having established both the definition of simultaneity and the dependence of simultaneity upon the reference frame from which clocks are observed, Einstein is now ready, in Section 3, to derive the these transformations. Again, his assumptions will be the Principle of Relativity and the independence from the observer's reference frame of the speed of light.

The reader is once more asked to imagine two reference frames, precisely those of Section 2, but described more specifically this time. The stationary frame, K, has co-ordinates x, y, z, and t. The moving reference frame, k, with relative velocity v in the positive direction along the X-axis, has co-ordinates ξ, η, ζ, and τ. Einstein observes that a vector (x,y,z,t) defines an event in the stationary reference frame -- specifically, the event at the point defined by those space co-ordinates that occurs at time t. The same event, observed from the moving frame, would be defined by the vector (ξ,η,ζ,τ), occurring at a corresponding point in that frame. at time (by clocks synchronized in the moving frame) τ.

"Our task," says Einstein, "is now to find the system of equations connecting these quantities" -- that is, the relation between the two sets of co-ordinates for the same event. He begins by seeking an expression for the τ of an event as a function of the space and time co-ordinates of that event in the stationary system.

He claims that, "it is clear that the equations must be *linear* on account of the properties of homogeneity which we attribute to space and time." In the present account, we will leave aside the question of how this clarity is attained by the reader seeking to understand the argument, and simply assume it.

We are now ready for the critical step. As in Section 2, we are to imagine that the two systems have the same origin at time τ_0 in the moving system, which corresponds to time 0 in the stationary system.[3] For convenience, we replace the variable x in the stationary system by

[3] Einstein set $t_0 = t$; I set $t_0 = 0$, simplifying the algebra.

$x' = x - vt$. If a point is at rest in the moving system, then its X-co-ordinate will equal $x_0 + vt$ for all t. Hence for such a point at rest, $x' = x_0$ will be a constant, independent of time.

At time $\tau_0 = t_0 = 0$, a ray is emitted from the origin along the positive X-axis. At time τ_1, when it has reached the point whose X-co-ordinate is x' in the stationary system, it is reflected back along the axis. And at time τ_2 it arrives back at the origin (i.e., at the origin of the moving system, where $x' = \xi = 0$). By the definition of simultaneity, "we then," says Einstein, "must have $1/2(\tau_0 + \tau_2) = \tau_1$, or, by inserting the arguments of the function τ and applying the principle of the constancy of the velocity of light in the stationary system:"

$$\frac{1}{2}[\tau(0,0,0,0) + \tau(0,0,0,\frac{x'}{c-v} + \frac{x'}{c+v})] = \tau(x',0,0,\frac{x'}{c-v})$$

The simultaneity equation makes explicit that the emission event has co-ordinates $(0,0,0,0)$ in the stationary system; the reflection event has co-ordinates $(0,0,0,\frac{x'}{c-v} + \frac{x'}{c+v})$ in the stationary system; and the return event has co-ordinates $(x',0,0,\frac{x'}{c-v})$ in the stationary system.

The above equation, plus one more that will be discussed in a moment, provides the basis for the derivation of the Lorentz transformations by purely mathematical means. Hence if we understand the rationale for this equation, we understand, in some sense, the Lorentz transformations that are the mathematical core of Special Relativity.

Since Einstein now uses a calculus step in his derivation, I will digress for a moment to show how it can be carried out algebraically. Since τ has been assumed to be a linear function of x and t, hence of x' and t, we may write it as $\tau = at + bx'$. Then, from the previous equation of simultaneity, $\tau_0 = 0$; $\tau_1 = a\{x'/(c-v)\} + bx'$; and $\tau_2 = a\{x'/(c-v) + x'/(c+v)\}$. Solving the latter two equations simultaneously, we can find a in terms of b:

$$a = \frac{-bv}{c^2 - v^2}$$

Replacing a by this value in the linear equation for τ, we obtain:

$$\tau = a(t - \frac{v}{c^2 - v^2}x')$$

This is the equation at the top of page 45 of the relativity paper. Since the velocity of light is c, the equation for the transformation of ξ can be obtained immediately by multiplying both sides of this equation by c. There remains the task of finding the transformations of the η

and ζ coordinates, perpendicular to ξ. Here Einstein (bottom of page 44) asks the reader, rather cryptically, to imagine "an analogous consideration ... applied to the axes of Y and Z ... it being borne in mind that light is always propagated along these axes, when viewed from the stationary system, with the velocity $(c^2 - v^2)^{1/2}$."

What the reader is being asked to imagine here is an axis of the moving co-ordinate system perpendicular to the direction of its motion. Suppose an observer in the moving system sees a light ray go a distance $c\tau$ along the Y-axis, perpendicular to X. Then, since the axis will have moved a distance, $v\tau$, in the same time, in the direction of the X-axis, but parallel to itself, an observer in the stationary system will see the same light ray take a diagonal path of length ct from the initial origin of the perpendicular axis to the point of reflection, and a diagonal path back to the new location of the origin. By the Pythagorean Theorem, this implies that, viewed from the stationary system, the Y-component of the velocity of the ray will be $(c^2 - v^2)$. Moreover, since the velocity will be the same in both directions, the times to and from the point of reflection will also be the same. Constructing the events for the times when the ray is emitted, is reflected, and returns to its starting point, we get the equation:

$$\frac{1}{2}[\tau(0,0,0,0) + \tau(0,0,0,\frac{2y}{(c^2 - v^2)^{1/2}})] = \tau(0,y,0,\frac{y}{(c^2 - v^2)^{1/2}})$$

Solving this equation as we did the previous one, we find, consistent with Einstein's conclusion, that τ is independent of y, and by symmetry, also of z.

Our sketch of Einstein's derivation of the Lorentz transformation shows that, as we claimed, the result follows by the application of elementary algebra to initial equations that are motivated by the simple thought experiments, involving mental imaging, described in the previous section of this paper.

Discussion

It is time now to draw some general conclusions from our examples about the nature of understanding. First, I will characterize the understanding process as it exhibits itself in those examples. Next, I will have a few words to say about another understanding process that is visible elsewhere in the relativity paper. Then, I will discuss some artificial intelligence programs that illustrate a concept of understanding similar to, if not identical with, the one

illustrated by our examples. I will conclude with comments on some of the implications of our findings for the theory of human cognition.

The Process of Explanation

In each section of the relativity paper that we have examined, the process of explanation is the same. The author provides a description of a situation in natural language, and invites the reader to form a mental image of the situation that has been described. Then, by inspection of the mental image, the reader is to infer an equation. Finally, the author may manipulate algebraically the equations obtained in this way in order to derive other equations. Understanding is a two-step process: first there is a translation of the natural language into a "picture," then a translation of the picture into an equation.

Understanding Derived Equations

Had we continued our examination of Einstein's relativity arguments through the remainder of his paper, we would have also observed a "reverse" understanding process: the interpretation of a derived equation by means of a mental image. For example, on page 48, Einstein derives the equation of a moving rigid sphere viewed from a stationary reference frame. Immediately after giving the equation, he says:

> A rigid body which, measured in a state of rest, has the form of a sphere, therefore has in a state of motion -- viewed from the stationary system -- the form of an ellipsoid of revolution . . ."

The reader, clearly, is expected to recognize the equation as defining an ellipsoid of revolution. The explanation proceeds from equation to picture.

Comparison with AI Programs

Let us return to the examples of the previous sections, where the understanding proceeds from verbal description to picture to equation. The process here bears a close resemblance to the methods used in several artificial intelligence systems that have tried to capture the phenomenon of "understanding." In the UNDERSTAND program, for example, Hayes and Simon (1974) modeled the first half of the process. Problems were presented to the UNDERSTAND program in verbal form, and the program undertook to construct an internal model (image) of the problem situation and of the legal moves that could change that

image.

An even closer match can be found in Gordon Novak's ISAAC program (1977), which takes physics problems described in English; creates an internal model of the problem situation; then draws upon the information in the model to set up equations. The model is sufficiently image-like that Novak was able to construct an auxiliary component of ISAAC that draws a representation of the problem situation on a CRT.

Although ISAAC would have to be provided with additional knowledge to understand the relativity paper, it's basic structure would allow it to accommodate such knowledge. It now lacks knowledge that would allow it to construct reference frames and to represent bodies moving in them, as well as knowledge for translating these kinds of images into equations.

Implications for Psychology

The picture of explanation and the nature of understanding that emerges from our analysis of the relativity paper is presaged not only by the UNDERSTAND and ISAAC programs, but also by earlier research on human processes in solving algebra and physics problems (Paige and Simon, 1969; Simon and Simon, 1977; Larkin, 19XX). In the very simple tasks used in that research, the subjects -- especially the more skilled among them -- did not translate word problems directly into equations, but first converted them into mental images of the situations described, and then converted these images into equations. We seem to be dealing, therefore, with a very general set of processes, at least within the realm of problems relating to concrete situations in a physical world.

The images required for handling the kinds of problems that have been studied previously, both in the laboratory and in AI programs, are quite simple. If complexity is ever present, we would expect to find it in some problem area that had a reputation of profundity -- the theory of Special Relativity being a good candidate. What we find, however, in the thought experiments that Einstein proposes to his readers, is not great complexity, but situations that can be captured in relatively simple pictures.

Several features of Einstein's own descriptions of these situations deserve comment.

First, although the theory is concerned with frames of reference that move relative to each other, only very elementary aspects of motion have to be handled in the images. The motion is reduced to a sequence of, at most, three discrete events, each of which is to be described in terms of the co-ordinates in some reference frame. The visualization does not require a true dynamics, but only a short sequence of discrete static pictures.

Only in some global way does one reference frame have to be imaged in motion relative to another. By checking back on our examples, it can be verified that all of the detailed steps of "seeing" make reference to a single reference frame. The final simultaneity relation is constructed by relating, for each of three events, the co-ordinates of that event for each of two reference frames, where these co-ordinates have already been worked out separately. One proceeds by a careful process of divide and conquer, only a small part of the total situation being attended to at any one moment.

The idea that all motion is relative -- that there is no "preferred" stationary reference frame -- is central to the Theory of Relativity. It is ironic, therefore, that in every one of his examples that involve two reference frames, Einstein designates one of these frames as "stationary," and the other as moving relative to it. To be sure, he sometimes places "stationary" in quotes, as I have been doing, but he uses quotes only twice, while he omits them in more than half a dozen cases. This usage seems extraordinary in view of his strong introductory statement that

> [T]he phenomena of electrodynamics as well as of mechanicws possess no properties corresponding to the idea of absolute rest. . . . [T]he view here to be developed will not require an "absolutely stationary space" provided with special properties.

This apparent contradiction in Einstein's posture can be resolved, I think, in a rather straightforward way. Einstein constructs his pictures, we have said, as an invitation to readers to form a mental image, which image is to be used as a basis for understanding an equation. Now in the theory, there is no privileged reference frame; the relation between two frames in relative motion is wholly symmetric. But the readers, to form an image, must choose *some* reference system as their own viewing point. It does not matter which frame they choose; the story would have the same outcome whichever they selected. But they must

select one. No one can image the world from two reference planes simultaneously.

It might be thought that this last assertion is contradicted, for example, by the ability people have of interpreting and understanding cubist paintings. But the contradiction is only apparent. In another place, I have discussed the related phenomena of the reversible Necker Cube and so-called "impossible figures" (Simon 1967). There I showed that mutually incompatible views of the same display are processed by "time sharing" -- that is by alternating between two images, not by holding them simultaneously.

We conclude that when Einstein refers to a reference frame as "stationary," he means "stationary relative to the viewer who is invited to form an image of the situation." The frame is not stationary in any more absolute sense. It is certainly not privileged with respect to the theory that is constructed, in which all the laws are invariant under Lorentz transformations of the reference frame.

How Complex is a Mental Image?

Einstein can hardly be called prolix in his initial presentation of the Theory of Relativity. Nor can we suppose that he was writing down to his readers. Hence the relative simplicity of the images he asks his readers to construct is striking. Evidently, even at the highest reaches of physics, the human mind takes small steps and operates with simple pictures containing limited information.

The images of the paper on relativity can perhaps provide us with at least some estimates of the upper bounds of human imaging capability -- of the extent to which problems have to be factored into their component parts before the human mind can encompass them. Our analysis of these images -- leading to the recognition of their relative simplicity, and of the way they are used to understand and motivate the theory -- reinforces the belief that we do indeed have to divide, and divide repeatedly, in order to conquer complexity.